University of Groningen

The nature of the beast. Analyzing and modeling computer network traffic.

Stegeman, Albertus Wilhelmus

Publication date:
2002

Summary

Since the beginning of the 1990s a growing number of computer networks from all over the world has been linked together. Nowadays, they constitute the global network called the *Internet.* The Internet has hundreds of millions of users, from private individuals and corporations to government officials and scientists, and offers a broad range of applications: *e-mail*, *newsgroups*, *remote login*, *file transfer*, *audio and video streams*, *chatrooms* and the popular *World Wide Web*. Although the Internet and computer networks in general have many benefits, some technological difficulties have been encountered. As surely everyone has experienced, the operation of transferring data between two computers in a network can be quite troublesome. At busy hours, when there are a lot of active users, the network can be congested, resulting in long transmission delays or difficulties in establishing a connection between the source and destination computer.

In order to get a better understanding of the dynamics of the data traffic in computer networks, a number of empirical studies of network traffic measurements has been conducted. As a benchmark for comparison voice traffic in the traditional telephone system has been used. In the telephone system call arrivals can be modeled by a Poisson process, i.e. with exponential inter-arrival times. Moreover, the distribution of call lengths has an exponentially bounded tail. From an engineering perspective these are very convenient properties, since the long-term arrival rate of the Poisson process, together with the mean call length, roughly determines the capacity of the network that guarantees reliable telephone communication.

Measurements of computer network traffic

The situation in computer networks, however, has been found to be very different. Computer network traffic has been studied at two levels: the *application level* and the *packet level.* At the application level file sizes, connection durations and transmission times are the main subjects of analysis. Instead of exponential their distributions appear to be *heavy-tailed*, i.e. $\mathrm{P}(X > x) \sim c x^{-\alpha}$, $x \to \infty$, with $c > 0$ and $\alpha \in (1, 2)$. Consequently, $\mathrm{Var}(X) = \infty$ and extremely large values of X occur with non-negligible probability. At the packet level the so-called *workload* of the network is measured. When a file is sent from a source to a destination computer, it is decomposed into small *packets* which are sent through

the network cables. After arriving at the destination computer, the packets are put together again and the original file is reconstructed. The workload on a cable or link in the network is measured by counting the number of packets or bytes passing the measurement point in a small time interval, e.g. 1 second. In this way, the workload per second is determined. Even for large time intervals, the dependence in the series of workload measurements appears to be rather strong: at large lags the sample autocorrelations still seem significantly different from zero. This phenomenon is often referred to as *long-range dependence.* This would indicate that random cycles of arbitrary length are present in the workload data. Another striking feature is that when the time interval used for measuring the workload is increased, the relative variability of the workload remains roughly the same, or, in other words, the workload shows a similar burstiness across a wide range of time scales. This property has been observed for time intervals ranging from 0.01 up to 100 seconds and resembles, in some sense, the theoretical notion of distributional *self-similarity.* Unlike voice traffic, computer network traffic does not smooth out when viewed at increasingly larger time scales.

Heavy tails, long-range dependence and self-similarity are believed to be present in traffic measurements on networks of different scales providing different applications, from the late 1980s till the present day. Therefore, these three features are regarded as *traffic invariants.* On the whole, this implies that computer network traffic behaves rather erratic compared to voice traffic in the telephone system, and, hence, that computer networks are a great challenge to the engineer and the scientist.

Non-stationarity versus long-range dependence

By definition a stochastic process exhibiting long-range dependence is stationary, i.e. its underlying distribution does not change during the time the process is observed. However, no general test for the stationarity of an observed time series is available. Also, the graphical methods that are often used to detect long-range dependence in a time series are not very reliable. It is well-known that these graphical methods can interpret non-stationarities like shifts in the mean or a slowly decaying trend as the presence of long-range dependence. In Chapter 3 of this thesis we show that a realization from a non-stationary ARIMA(p,1,q) process, with appropriate parameter values, can indeed exhibit long-range dependence in this 'graphical' sense. In Chapter 4 we analyze series of workload measurements in various computer networks and find that most of them can be modeled by an ARIMA(p,1,q) process, with small p and q. This shows that, when using graphical methods, it is virtually impossible to distinguish between non-stationarity and long-range dependence in an observed time series. Here however, given the complicated nature of a computer network, with applications and connections being activated and terminated during the measurement period, the option of non-stationarity is probably the most reasonable one.

Modeling computer network traffic

An attempt to give a 'physical' explanation for the observed traffic characteristics has been made by using a mathematical modeling approach. Two simple models have been proposed which both use the assumption of heavy-tailed transmission times to explain the long-range dependence in the workload of the network. Also, it is shown that the centered and properly normalized cumulative workload can be approximated, in some sense, by a self-similar process. One of these two models, the *ON/OFF model* introduced by Willinger et al. [106], is the subject of Chapters 5 and 6 of this thesis. In this model, traffic is generated by M independent and identically distributed *ON/OFF sources.* If a source is ON it transmits data at unit rate, e.g. 1 byte per time unit. If it is OFF it remains silent. In this way, every individual ON/OFF source generates a binary *ON/OFF process.* The lengths of periods in which a source is ON, the *ON-periods*, are independently drawn from a heavy-tailed distribution. Analogously, the *OFF-periods* are also heavy-tailed. The sequences of ON- and OFF-periods are assumed independent. It has be shown by Heath et al. [45] that the stationary version of the ON/OFF process of an individual source exhibits long-range dependence. Using independence, the same is true for the total workload, i.e. the superposition of the M ON/OFF processes.

In Willinger et al. [106] it is shown that the centered cumulative workload up to time T, when properly normalized, converges in finite dimensional distributions to fractional Brownian motion if first $M \to \infty$ and then $T \to \infty$. In Taqqu et al. [101] the limits are reversed and a different normalization is used to obtain stable Lévy motion as limit process. Both fractional Brownian motion and stable Lévy motion are self-similar, but their dependence structures are totally different. The increment sequence, at equidistant instants of time, of fractional Brownian motion is stationary and exhibits long-range dependence (thus preserving the long-range dependence in the pre-limit workload), while the increments of stable Lévy motion are independent. In Chapter 5 of this thesis we consider simultaneous limit regimes in which M is a non-decreasing function of T, converging to infinity as $T \to \infty$. We show that when M grows faster than some 'critical rate' fractional Brownian motion is obtained in the limit. On the other hand, if M grows slower than this 'critical rate' stable Lévy motion appears as limit process.

In Chapter 6 we use the framework of the ON/OFF model to study the number of ON-periods up to time T exceeding a high threshold. Again, we consider simultaneous limits of M and T. Moreover, also the threshold depends on T. We distinguish between the 'slow' and 'fast' growth conditions on M. Although different approaches are needed, in both cases we are able to show that the number of exceedances converges to a Poisson random variable if the threshold satisfies a balancing condition guaranteeing a constant average number of exceedances in the limit. We also show that if the threshold grows slower than this 'balancing rate', the number of exceedances satisfies the Central Limit Theorem.